ESTIMATING PARAMETERS IN A FINITE MIXTURE OF PROBABILITY DENSITIES

Richard P. Heydorn and Rehka Basu*

ABSTRACT

The problem of estimating the number of components, and the other parameters, in a finite mixture of probability densities is formulated as a continuous mixture estimation problem. Representing the finite mixture as $h = \int f(.,\theta)dG(\theta)$, where G changes only on a finite number of points, it is shown that under fairly general conditions it is possible to represent this mixture as $h = \int f(.,\theta)g(\theta)d\theta + \delta$ where the error δ can be made arbitrarily small by properly choosing g. It is further shown that this second representation is consistent with the first one, by showing that the c.d.f. of g can be made to converge weakly to the true G. Estimators are proposed for g based on kernel estimators and linear programming methods.

Keywords: Integral equation; Weak convergence; Linear programming; Graphical method; Spline.

---

*Richard P. Heydorn is with the Johnson Space Center, Houston Texas and Rehka Basu is with the Honeywell Electro-Optics Division, Lexington Massachusetts.

(NASA-TM-101783) ESTIMATING PARAMETERS IN A FINITE MIXTURE OF PROBABILITY DENSITIES (NASA. Lyndon B. Johnson Space Center) 22 p

N89-71032

Unclas

00/65 0213159

1. INTRODUCTION

Let X be a random variable whose observations $x \in \mathbb{R}$ ($\mathbb{R}$ denotes the real line) are distributed according to the mixture density

$$h(x) = \sum \lambda_j f(x, \theta_j) \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The components of this mixture, $f(\cdot, \theta_j)$, $j = 1, 2, \ldots, N$ are each members of a given family of densities, $\mathcal{F}$, parameterized according to the real variable θ. The mixing proportions $\lambda_j \in \mathbb{R}$, $j = 1, 2, \ldots, N$ are nonnegative and $\Sigma \lambda_j = 1$.

Finite mixture distributions have turned out to be a suitable model for a large number of physical processes. In their recent book, Titterington, Smith, and Makov (1985) mention a number of applications of the finite mixture model. Our interest in the model arose from a problem in remote sensing in which it was required to estimate the proportion (i.e., a mixing proportion in equation(1)) of an agricultural area that was planted to a given crop from satellite derived spectrometer measurements of the area. In this problem not only were the mixing proportions of equation (1) unknown but the other parameters N, θ_j, $j = 1, 2, \ldots, N$ were also unknown.

In the case where N is known, several authors have proposed estimation methods for determining the remaining parameters. Among these is the class of maximum likelihood methods which contains the EM algorithm of Dempster, Laird, and Rubin (1977). It is known (Zacks 1971) that when the model is specified, in this case N being given, the maximum likelihood estimator, under rather weak conditions, is consistent, is asymptotically unbiased, and is an asymptotically minimum variance estimator. However, when N is unknown these properties need not hold. Holding the number of samples observations fixed, Redner, Kitagawa, and Coberly (1981) point out that the likelihood function evaluated at the maximum likelihood estimate $(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_N, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_N)$ increases as N increases. This behavior has prompted the introduction of penalized maximum likelihood estimators. Redner, Kitagawa, and Coberly (1981), for example, applied such a penalized estimator based on the so-called AIC, which was originally introduced by Akaike (1973, 1977).

If N is believed to be a certain value, say N_0, then one could presumably test the null hypothesis that $N=N_0$ against some alternative. There have been a number of proposed statistical tests of this kind. The book by Titterington, Smith, and Makov (1958) has a good summary of this approach.

Our approach for determining N, as well as the other parameters in the mixture, uses an integral equation representation of the finite mixture. This representation has the form

$$h(x) = \int f(x,\theta)g(\theta)d\theta + \delta(x) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where the error function, δ, can be made small (in norm) by an appropiate choice for g. Given h, the function g can be chosen so that it is a density function having N modes and high density on small intervals that contain the true parameter values $\theta_1, \theta_2, \ldots, \theta_N$.

Moveover, the cumulative distribution related to g will approximate a distribution with jumps at these θ-parameter values and whose jump sizes are $\lambda_1, \lambda_2, \ldots, \lambda_N$. Thus, the determination of $N, \lambda_j, \theta_j, j=1, 2, \ldots, N$ depend only on the determination of a single function, g. When h is unknown, but can be estimated from the data, then the problem is to estimate g by methods which tend to preserve the above desired characteristics.

This approach bears some similarity to the one proposed by Medgyessy (1977) in which a test function is derived from the mixture and the graph of the test function is more infromative about the mixture parameter values than is the mixture itself. In particular the graph of the test function should have N narrow peaks that are much more pronounced than any peaks that may be present in h. The structure of the test function thus reveals the number of components in the mixture and may reveal other information concerning the other parameters. In our case g would be representative of Medgyessy's test function.

We will begin our discussion of finite mixtures by first considering the case where the parameter, θ, is a translation parameter, and then we will consider the more general situation in which θ is not necessarily a translation parameter. In both cases certain convergence results will be presented which will establish the consistency between the finite mixture and the integral equation representation. We will conclude by proposing estimators of the g-function and examine some of the properties of these estimators using numerical simulations.

2. INTEGRAL EQUATION REPRESENTATION

We can write the finite mixture of equation (1) as

$$h(x) = \int f(x,\theta)dG(\theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

where G is a cumulative distribution function (c.d.f.) that is a step function with its discontinuities at the points $\theta_1, \theta_2, \ldots, \theta_N$ and where $G(\theta)=0$ for $\theta<\theta_1$ and $G(\theta)=1$ for $\theta>\theta_N$. We will assume that f is a continuous function on $(-\infty,\infty)\times A$ where A is a closed interval containing the set $\{\theta_1, \theta_2, \ldots, \theta_N\}$ and that $f(.,\theta)$ is a density function for any $\theta\in A$.

The family $\mathbf{G}$ of all c.d.f.'s G induces, through the kernel f, a family of mixtures $\mathcal{H}$. If given $h\in\mathcal{H}$, there exists a unique $G\in\mathbf{G}$, then $\mathcal{H}$ will be called identifiable (Teicher 1963). In this paper we will deal only with identifiable families.

We begin by considering the case where θ is a translation parameter. In this case $\mathcal{H}$ is the family of all mixtures that can be generated from translates of some given density f. Any $h\in\mathcal{H}$ has the form

$$h(x) = \int f(x-\theta)dG(\theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Yakowitz and Spragins (1963) prove that $\mathcal{H}$ is identifiable; however, it is interesting to note that identifiability for this case can be easily deduced from the Caratheodory trigonometric moment theorem (Grenander and Szego 1958, pages 56-61). Indeed, letting $\mathbf{h}$, $\mathbf{f}$ denote the Fourier transform of h,f respectively, it follows from equation (4) that $\mathbf{h}(\omega)/\mathbf{f}(\omega) = \Sigma\lambda_j \exp(i\omega\theta_j)$ where the summation goes from 1 to N. By the Caratheodory theorem this is a unique representation for any collection of values ω_k, $k=1, 2, \ldots, N$ for which $\mathbf{f}(\omega_k)\neq 0$.

To obtain the integral equation representation for this case, we convolve h in equation

(4) with a density function t concentrated on an interval $(-\epsilon,\epsilon)$, $\epsilon>0$, to obtain

$$(h*t)(x) = \int f(x-z)g_t(z)dz \quad \ldots\ldots\ldots\ldots\ldots\ldots(5)$$

where

$$g_t(z) = \int t(z-\theta)dG(\theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots(6)$$

Thus

$$h(x) = \int f(x-\theta)g_t(\theta)d\theta + \delta(x) \quad \ldots\ldots\ldots\ldots\ldots(7)$$

with $\delta(x) = h(x) - (h*t)(x)$. Letting $||.||$ denote the supremem norm, and noting that h is a continuous density,

$$||\delta|| \leqq \int \sup_x |\, h(x)-h(x-z)|\, t(z)\, dz \to 0$$

as $\epsilon \to 0$. Also for any $q \in \mathbb{C}(A)$ (where $\mathbb{C}(A)$ is the set of all continuous functions on A) we have from equation (6),

$$\int q(z)g_t(z)dz = \int [\int q(z)t(z-\theta)dz]\, dG(\theta) \to \int q(\theta)dG(\theta)$$

as $\epsilon \to 0$. This implies that the c.d.f. G_t of g_t converges weakly to G.

Since $||\delta|| \to 0$ and since $G_t \to G$ weakly, we will say that our integral equation representation, equation (7), is consistent with the finite mixture representation given by equation (4). It follows also that

$$g_t(z) = \sum \lambda_j\, t(z-\theta_j) \quad \ldots\ldots\ldots\ldots\ldots\ldots(8)$$

and so g_t is a mixture density with the same parameter values as h; but, by choosing t to be a unimodal density concentrated at 0, the graph of g_t is generally more revealing than is the graph of h. In fact h can be unimodal while g_t can have N modes at the points θ_1, $\theta_2, \ldots, \theta_N$.

In the general case where θ is not necessarily a translation parameter, we can proceed in the following way. Consider the functions $t(.,\theta_j)$ $j = 1,2,\ldots,N$ where each $t(.,\theta_j)$ is a density function concentrated on a small interval that contains the true parameter value θ_j in its interior. Define

$$h_N(x) = \sum \int f(x,\theta)\, t(\theta,\theta_j)\, d\theta \quad \ldots\ldots\ldots\ldots\ldots\ldots(9)$$

Since h can be written as

$$h(x) = \sum \lambda_j \int f(x,\theta_j)\, t(\theta,\theta_j) d\theta$$

we have

$$||h-h_N|| \leqq \sum \lambda_j \int \sup_x |f(x,\theta) - f(x,\theta_j)|\, t(\theta,\theta_j) d\theta$$

Since $f(.,\theta)$ is a continuous density function for all $\theta \epsilon A$, we can make $||h-h_N||$ arbitrarly small by forcing each density function $t(.,\theta_j)$ to be increasingly more concentrated on θ_j, i.e., letting the interval support of $t(.,\theta_j)$ approach zero in size.

Since the parameter values θ_j, $j=1,2,\ldots,N$ are not known, we can not choose the $t(.,\theta_j)$-functions directly. Instead, we will consider a sequence of normalized B-spline functions of order 2, or greater, defined over equally spaced knots that partition the interval A. Since the support for the splines overlap, any given θ_j will be contained in the interior of the support for some B-spline function. For such a set of B-splines, $\{B_k, k=1,2,\dots,M\}$ define

$$h_M(x) = \sum \alpha_k \int f(x,\theta) B_k(\theta)\, d\theta$$

By letting $M \to \infty$ it is clear, that for this case, we will also have $||h-h_M|| \to 0$, provided of course that for each choice of M we pick the α_k-values properly. In the next section we will propose a linear programming approach for computing these α_k-values

By letting

$$g_M(\theta) = \sum \alpha_k B_k(\theta) \quad \text{......................}(10)$$

where the summation goes from 1 to M, we have

$$h(x) = \int f(x,\theta)\, g_M(\theta)\, d\theta + \delta_M(x) \quad \text{............}(11)$$

where $\delta_M = h-h_M$. To establish that this representation, in equation (11), is consistent with the one of equation (3), it remains to show that the c.d.f. of the density g_M weakly converges to G.

Theorem. Let $\mathcal{H}$ be identifiable and let h be defined according to equation (3). For M = 1,2, ... define G_M to be the c.d.f. corresponding to g_M (where g_M is defined according to equation (10)). Define h_M as

$$h_M(x) = \int_A f(x,\theta)\, g_M(\theta)\, d\theta$$

If $||h-h_M|| \to 0$ $(M \to \infty)$ then $G_M \to G$ weakly.

Proof. Denote a linear functional on $\mathbf{C}(A)$ as $\mathcal{L}$. When $\mathcal{L}$ is associated with some given c.d.f., F, denote it by $\ell(.,F)$ (i.e., $\ell(.,F) = \int . dF$). First note that $\mathbf{C}(A)$ is separable and (e.g., from the Riez representation theorm) $||\ell(.,G_M)|| = 1$ for all M. Thus (Kolmogorov and Formin 1957, p. 94) the sequence $(\ell(.,G_M))$, or any subsequence of that sequence, contains a weakly convergent subsequence. Thus we can find a subsequence (M') so that $\ell(q,G_{M'}) \to \ell(q)$ for any $q \in \mathbf{C}(A)$. By the Riez representation theorem $\ell(q) = \int q\, dF$ and since $\ell(1,G_{M'}) = 1$ for all M', F is a c.d.f. (Feller 1971, p. 251). Since for any x, $f(x,.) \in \mathbf{C}(A)$ and $\ell(f(x,.),G_{M'}) \to \ell(f(x,.),G)$, by the identifiability of $\mathcal{H}$ it must be that F = G. To complete the proof simply note that the above argument implies that $\limsup \ell(q,G_M)) = \liminf(\ell(q,G_M)) = \ell(q,G)$.

3. ESTIMATORS OF g

Given the finite mixture representation

$$h(x) = \int f(x,\theta)g(\theta)d\theta + \delta(x) \quad \text{..................} \ (2')$$

we now turn to the problem of estimating g from an iid sample $X_1, X_2, \dots, X_n$ where each X_i is distributed according to the mixture density h.

Our primary purpose for estimating g is to obtain an estimate of N in equation (1). As discussed in the Introduction, our estimate of N will be based on the number of "pronouned" peaks in the estimated graph of g. Also, with reference to equation (1), the location of each peak (the mode) can be taken as an estimate of a θ_j-value; and, the size of each jump in the c.d.f. corresponding to the estimated g about a θ_j-value can be taken as an estimate of the λ_j-value. All of these estimates are subjective and depent upon a visual inspection of the density or c.d.f. graph.

A possible approach for estimating g would be to first estimate h, from the given sample, $X_1, X_2, \ldots, X_n$, and then, calling $\hat{\mathbf{h}}$ our estimate of h, minimize $\| \hat{\mathbf{h}} - \int f(.,\theta)\, g(\theta)\, d\theta \|_k$ with respect to g. Here $\|.\|_k$ is some norm, which may not be the supremum norm considered earlier. Since the family $\{f(.,\theta);\ \theta \in A\}$ is equicontinuous and uniformly bounded, the operator induced by the kernel f is compact (Lusternik and Sobolev 1961). The inverse operator, if it exists, will therefore be discontinuous. This means that any numerical solution for g can be very sensitive to small changes in $\hat{\mathbf{h}}$. In Tikhonov's regularization method (Tikhonov and Arsenin 1977), this discontinuity problem is addressed by appropiately constraining the solution for g by a linear combination of the L_2 norms of g and its derivatives.

Rather than use the general representation of equation (2') we will use the specific representations of equations (7) and (11). That is, we will estimate g_t for a suitable choice of t and g_M for a suitable choice of a spline series. In the case of g_t it is possibe to avoid the above mentioned discontinuity problem by deriving an operator from f that is continuous and, upon operating on h, will give g_t. We now discuss this approach.

Rewriting equation (5) in terms of Fourier transforms gives at a point ω

$$\mathbf{g}_t(\omega) = (\mathbf{t}(\omega)/\mathbf{f}(\omega))\,\mathbf{h}(\omega) \quad \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

Let $\mathbf{t}_f = \mathbf{t}/\mathbf{f}$. Then (assuming $\mathbf{t}_f$ is integrable)

$$g_t(\theta) = \int t_f(\theta - x)\, h(x)\, dx$$

Given our above random sample $X_1, X_2, \ldots, X_n$ an (unbiased) estimator for g_t is

$$\hat{g}_t(\theta) = (1/n) \sum t_f(\theta - X_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots.(13)$$

In the next secion we will consider this estimator for the case where f is a gamma density function i.e. $f(x) = (1/m!\gamma^{m+1})\, x^m \exp(-x/\gamma)$ for $x \geqq 0$ and $f(x)=0$ for $x<0$. For this case equation (12) becomes

$$\mathbf{g}_t(\omega) = ((1+i\omega\gamma)^{m+1}\mathbf{t}(\omega))\ \mathbf{h}(\omega)$$

If we choose t so that it has m+1 derivatives and the m derivatives as well as t vanish at $-\infty$ and ∞, then

$$t_f(x) = (1 + \sum ((m+1)!/k!(m+1-k)!)\ \gamma^k D^k)\ t(x) \ldots..(14)$$

where the summation is over k from 1 to m+1 and where D is the standard derivative operator ($D = d/dx$). Notice that the gamma function leads to an inverse operator that is a linear combination of derivative operators, D^k, $k=1,2, \ldots, m+1$ and is therefore discontinuous. The role of t, in this case, is to transform this discontinuous operator to a continuous one.

For the more general case where θ is not necessarily a translation parameter we turn to the problem of estimating g_M as given by equation (10). In this case we propose to minimize $||\hat{\mathbf{h}} - \int f(.,\theta) g_M(\theta) d\theta||_1$ with respect to g_M subject to certain constraints on g_M, which we will discuss in a moment. Here $||.||_1$ is the l_1 norm. Given the iid sample $X_1, X_1, \ldots, X_n$ (where recall each X_i is distributed according to the mixture h), $\hat{\mathbf{h}}$ is defined as the histogram, on the partition $x_0 < x_1 < \ldots < x_m$,

$$\hat{h}(x_j) = n_j / n$$

where n_j is the number of observations on $X_1, X_2, \ldots, X_n$ fall in $(x_{j-1}, x_j]$ for $j = 1, 2, \ldots, m$. The kernel, f, is also evaluated at these partition points.

This minimization will be expressed as a linear programming problem, viz,

$$\text{minimize: } \Delta_1 + \Delta_2 + \ldots + \Delta_m$$

$$\text{subject to: for } j = 1, 2, \ldots, m,\ k = 1, 2, \ldots, M,$$

$$-\Delta_j \leqq \hat{h}(x_j) - \int f(x_j, \theta) g_M(\theta) d\theta \leqq \Delta_j$$

$$\Delta_j \geqq 0,\ \alpha_k \geqq 0,\ \Sigma \alpha_k = 1$$

The constraints in this case force the resulting estimate of g_M, say $\hat{g}_M$, to be a density function. The solution of the linear programming problem is the α_k-values that define $\hat{g}_M$, which we will call $\alpha_1, \alpha_2, \ldots, \alpha_M$.

Since the sequence $(g_M(\theta))$ does not converge; but, instead the smoothed sequence $(\int q(\theta) g_M(\theta) d\theta)$ converges for all $q \in C(A)$, it would appear that we should also observe some smoothed graph of $\hat{g}_M$. In particular if (c.f. equation (10))

$$\hat{g}_M(\theta) = \sum \hat{\alpha}_k B_k(\theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots (15)$$

where B_k is a B-spline of order r then convolving $\hat{g}_M$ with a B-spline of order s would give a series expansion in terms of B-splines of order r+s with the same $\hat{\alpha}$-values. This is in fact what was done in the numerical studies discussed in the next section. The original expansion is in terms of second order B-splines and the smooth version is based on B-splines of order four.

Notice that this estimation approach is based on an l_1 norm as apposed to the supremum norm used in the theorem. Since h and h_M are uniformly continuous, and since the histogram will converge almost surely to h, it is easily seen that given $\epsilon>0$ it is possible to almost surely dominate $||h-h_M||$ by $||\hat{h} - \int f(.,\theta)g_M(\theta)d\theta||_1 + \epsilon$ for a fine enough partition $x_0<x_1<\ldots<x_m$.

4. NUMERICAL EXAMPLE

The first set of examples is based on a mixture of two translated gamma density functions of the form

$$h(x) = \sum(1/2)((x-\theta_j)/\gamma^2)(\exp-(x-\theta_j)/\gamma)\psi(x-\theta_j)$$

where $\psi(x)=1$ for $x\geqq 0$ and 0 otherwise and where $\gamma=.1$. The estimator of g_t, given in equation (13), was computed for a sample of size 1000 from this mixture. In equation (13) the function t_f was computed according to equation (14) taking the function t to be a fourth order B-spline with knots placed at intervals of length $\Delta=.1$. The function t_f therefore has the form

$$t_f(\theta) = t_4(\theta) + (2\gamma/\Delta)(t_3(\theta)-t_3(\theta-\Delta))$$

$$+ (\gamma^2/\Delta^2)(t_2(\theta)-2t_2(\theta-\Delta)+t_2(\theta-2\Delta))$$

where t_k, k=2,3,4 are second, third, and fourth order B-splines respectively. The expressions for these B-splines can be found in Schumaker (1981).

Figures 1-3 show the estimated mixture densities and the corresponding $\hat{g}_t$ estimated densities. In these figures h was estimated using a kernel estimator where the kernel was a third order B-spline with knots placed on intervals of length .1.

In the first figure the mixture consists of two components with $\theta_1 = .2$ and $\theta_2 = .3$. The spacing appears to be too close to resolve the components of the mixture by these methods as can be seen from the unimodal form of $\hat{g}_t$. As the components become more separated, as shown in Figures 2 and 3 the components show up. In Figure 2 $\hat{g}_t$ displays two modes even though the estimated h still appears to have no inflections. When $\theta_1=.2$ and $\theta_2=.4$, in Figure 3, the separation of the components can be seen in the mixture and is even more pronounced in the graph of $\hat{g}_t$.

The next set of examples is based on a mixture of two normal densities with different variances but the same means. In this case

$$h(x) = \sum(1/2)(1/(2\pi\sigma_j^2)^{1/2})\exp-((1/2)x^2/\sigma_j^2)$$

As indicated in the previous section a linear programming approach was used to estimate g_M. The actual form of this estimator is given in equation (15) with M=20.

Figures 4-6 show the histograms of h calculated from 1000 observations on this mixture with a cell size of .417 in the range of -5 to 5. In each case $\sigma_1^2=1$; but, $\sigma_2^2=2$

in Figure 4, $\sigma_2{}^2$=2.5 in Figure 5, and $\sigma_2{}^2$=3 in Figure 6. The coresponding graphs of $\hat{g}_M$ are also given in the figures. Twenty B-splines were used to compute $\hat{g}_M$ on equally spaced knots on the interval (.01,5).

The graph of $\hat{g}_M$, in Figure 4, suggests that two components may be present in the mixture;but, since the peaks are so disproportionate in size the smaller peak may be an artifact of noise. Figures 5 and 6, however, show that components whose variances differ by 1.5 or more could be distinguished.

Both sets of figures suggest that the graphs of $\hat{g}_t$ or $\hat{g}_M$ reveal the number of components in the mixture better than the other parameters. When components are not well separated any distinguishable peaks in the graphs of $\hat{g}_t$ and $\hat{g}_M$ tend to be disproportionate in size and therefore any estimate of the mixing proportions that could be derived by examining the c.d.f. of these functions would be considerably in error. The estimates of the location or scale parameters that could be derived by examining the location of the peaks are somewhat better.

5. CONCLUDING REMARKS

We have shown that in certain cases it is possibe to approximate a finite mixture as a continuous mixture, i.e., a mixture in which the mixing function is a continuous density function. We have further proposed methods for estimating this mixing function; and, from a few numerical examples have observed the behavor of the estimates. Our purpose was to develop a graphical approach for determining the number of components in the mixture and to a lesser extent to determine the other parameters in the mixture.

By approximating a finite mixture with a continuous mixture, one could possible loose uniqueness in certain cases. It is well known, for example that a finite mixture of normals in which the means and variances are allowed to vary is an identifiable mixture (Teicher 1963) or (Yakowitz and Spragins 1968). However, the same is not true of a continuous mixture of normals when both the means and variances vary, as was pointed out by Teicher (1960). The extent to which these multivariate cases limit our approach needs to be considered.

REFERENCES

Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle", Proc. Sec. Int. Symp. Inform. Theory, Budapest, pp 267-281 (1963)

Akaike, H., "On Entropy Maximazation Principle", P. R. Krishnaiah ed., Applications of Statistics, North Holland Amerstram (1977)

Dempster, A. P., Laird, N. M., Rubin, D. B., "Maximum Likelihood Estimation from Incomplete Data Via, the EM Algorithm", J. R. Statistical Soc. B 39, pp 1-38 (1977)

Feller, W., An Introduction to Probability Theory and Its Applications, Vol II John Wiley and Sons inc. (1971)

Grenander, U., Szego, G., Toeplitz forms and Their Applications, Berkeley, University of California Press, (1958)

Kolmogorov, A. N., Formin, S. V., Elements of the Theory of Functions and Functional Analysis, Vol I, Graylock Press (1957)

Lusternik, L. A., Sobolev, V. J., Elements of Functional Analysis, Frederick Ungar Pub. Co., (1961)

Medgyessy, P., Decomposition of Superpositions of Density Functions and Discrete

Distributions, John Wiley and Sons, (A. Halsted Press) (1977)

Redner, R. A., Kitagawa, G., Coberly, W. A., "Estimation of the Number of Components in a Mixture Model Using AIC", University of Tulsa, NAS-9-14689, (1981)

Schumaker, L. L., Spline Functions: Basic Theory, John Wiley and Sons, (1981)

Teicher, H., "On the Mixture of Distributions", Ann Math Stat, Vol 31, pp 55-73, (1960)

Teicher, H., "Identifiability of Finite Mixtures", Ann Math Stat Vol 34, pp 1256-1269 (1963)

Tikhonov, A. N., Arsenin, V. Y., Solutions of Ill-Posed Problems, V. H. Winston and Sons, (1977)

Titterington, D. M., Smith, A. F., Makov, U. E., Statistical Analysis of Finite Mixture Distributions, John Wiley and Sons (1985)

Yakowitz, S. J., Spragins, J. D., "On the Identifiability of Finite Mixtures", Vol 39, pp 209-214 (1968)

Zacks, S., The Theory of Statistical Inference, John Wiley and Sons, Inc. New York, (1971)

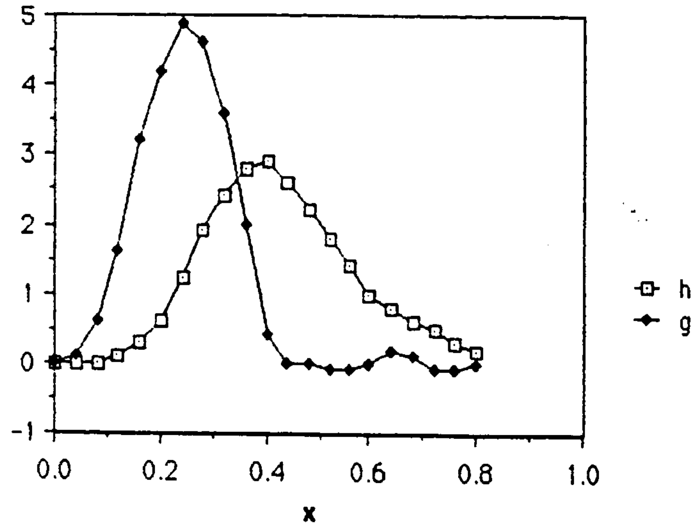

figure 1

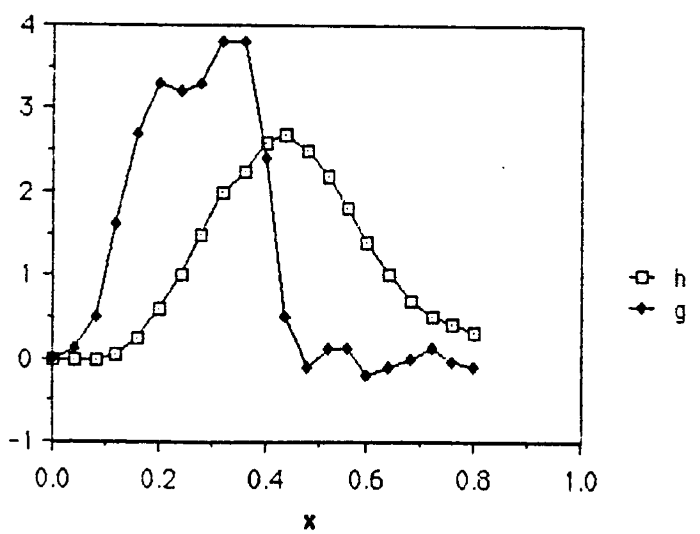

figure 2

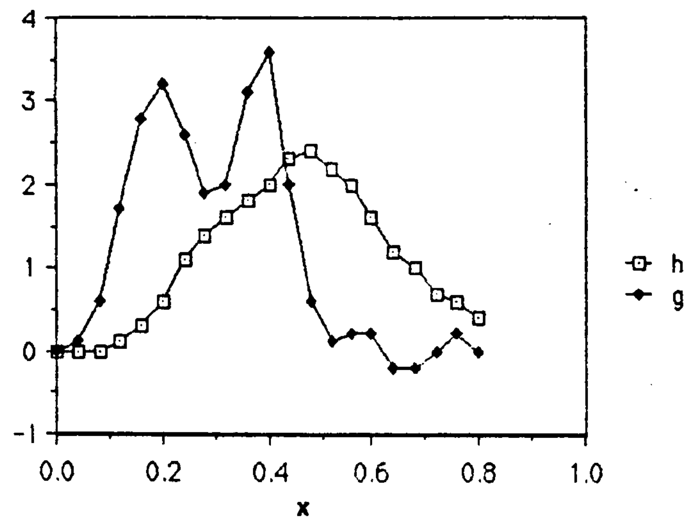

figure 3

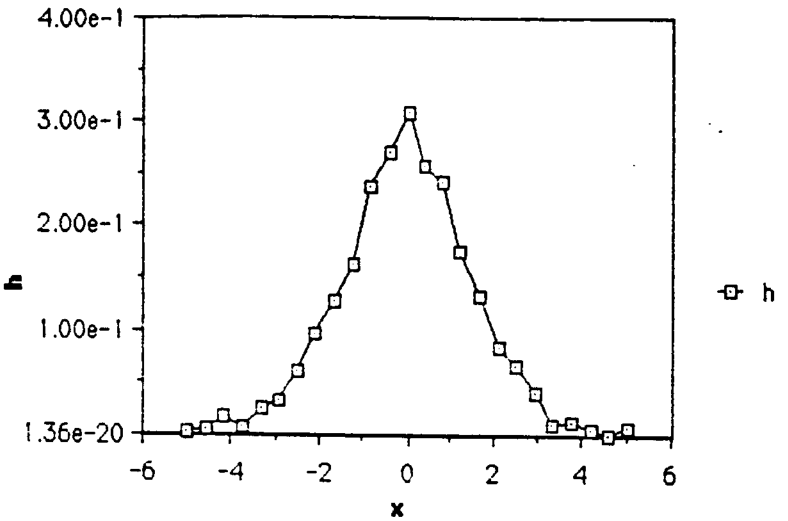

Figure 4 (a)

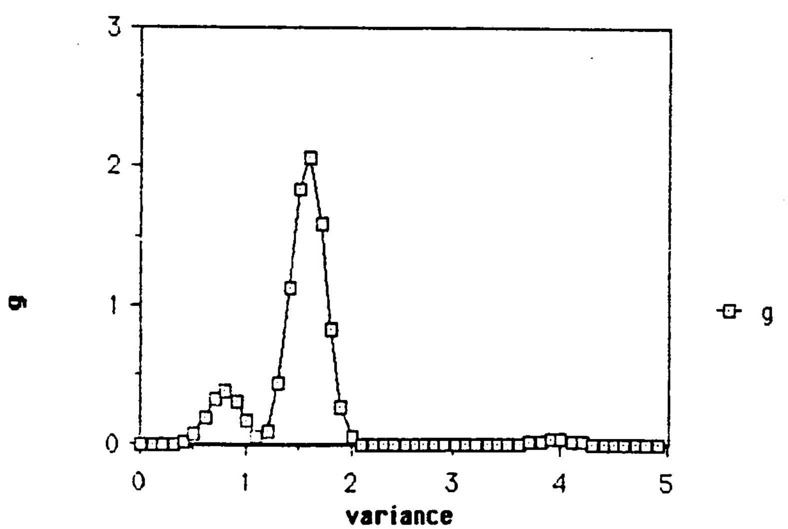

figure 4 (b)

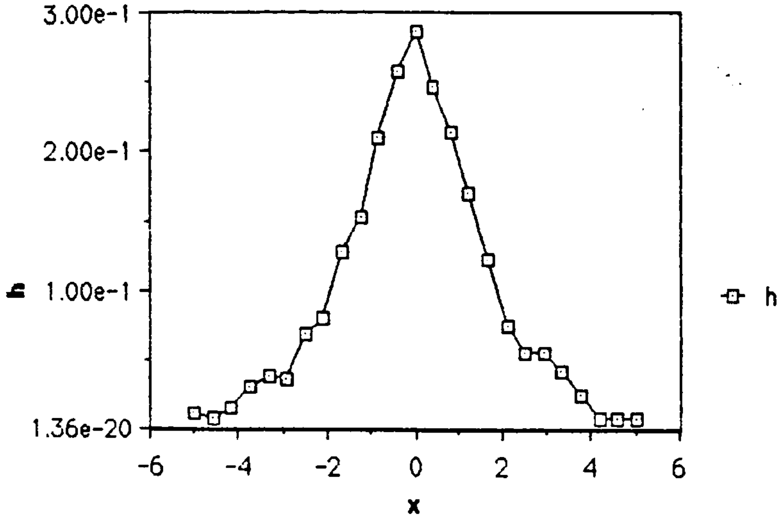

Figure 5 (a)

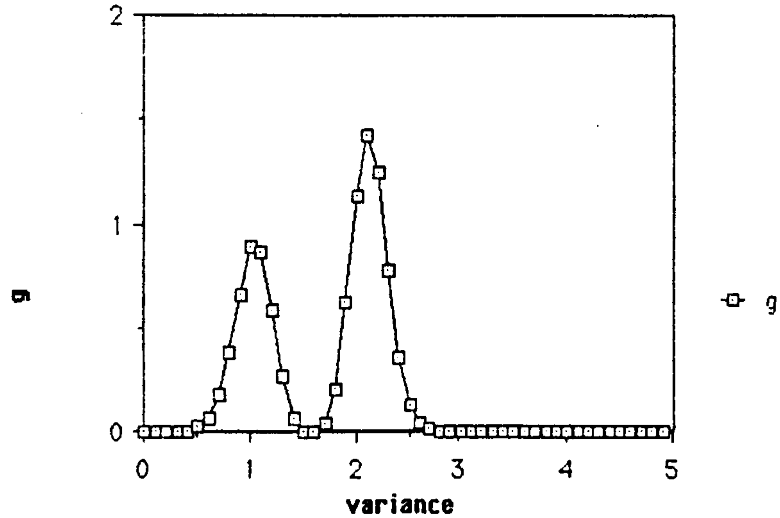

Figure 5 (b)

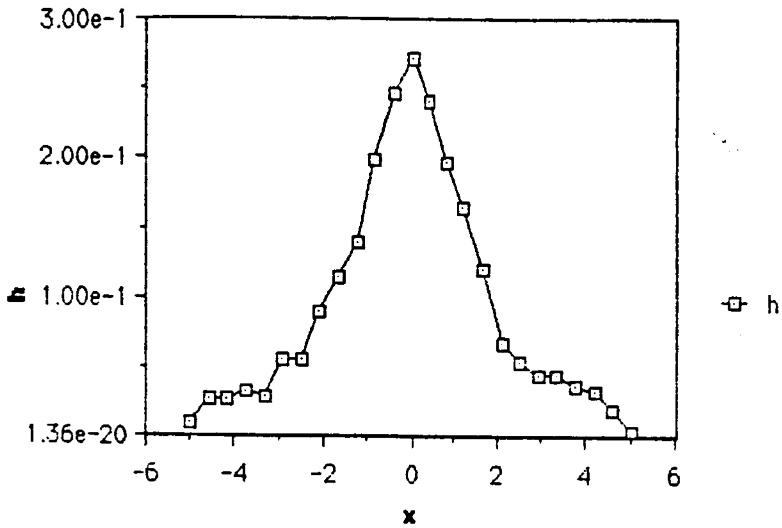

Figure 6 (a)

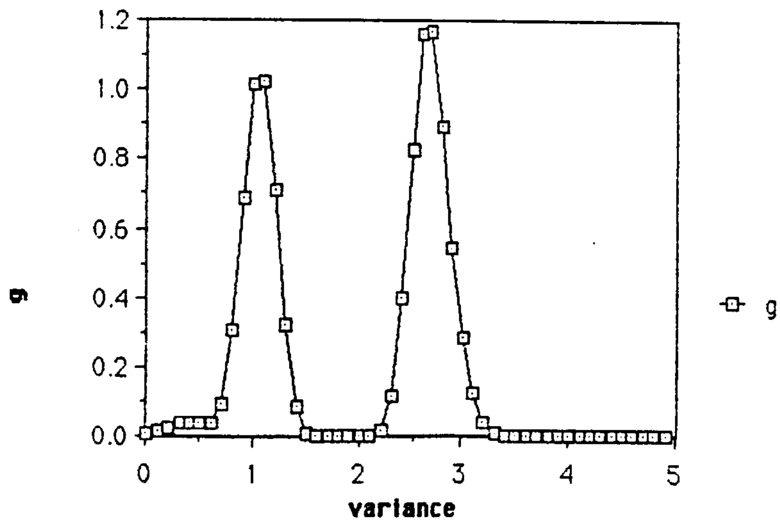

Figure 6 (b)